ABSTRACT

Customer Relationship Management (CRM) is an essential tool for the ongoing relationship between people, the suppliers and customers of goods and services. A company can proactively provide the best customer service possible while creating a database of customer preferences if every customer-company interaction could be recorded. Those transactions can be reviewed by all the departments including sales, marketing, and management and the information gathered by CRM efforts can be used to reduce costs and improve employee productivity. In this study, data mining tools are applied to the thermal hotels data in Marmara Region in Turkey. Data are taken from December to January and clustered to be able to benefit CRM applications.

KEYWORDS: Data Mining, CRM, Cluster Analysis, K-means, Tourism Marketing

INTRODUCTION

CRM uses information technology to create a cross-functional enterprise system that integrates and automates many of the customer serving processes in sales, marketing, and service that interact with a company’s customers. The business benefits of CRM are many. For example, CRM allows a business to identify and target their best customers, those who are the most profitable to the business, so they can be retained as lifelong customers for greater and more profitable services. It enables real-time customization and personalization of products and services based on customer wants, needs, buying habits, and lifecycle. CRM enables a company to provide a consistent customer experience and superior service and support across all the contact points a customer chooses. All of these benefits provide strategic business value to a company and major customer value to its customers. (O’Brien, p. 301-302-303)

CRM means presenting a single image of the company across all the many channels a customer may use to interact with the firm, and keeping a single image of the customer that is...
shared across the enterprise. CRM requires understanding who customers are and what they like and don’t like. It means anticipating their needs and addressing them proactively. It means recognizing when they are unhappy and doing something about it before they get fed up and churn which means go to a competitor. (Berry and Lionoff, p. 14)

CRM is fundamentally about the ongoing relationship between people, the suppliers and customers of goods and services. In theory, every customer-company interaction can be recorded, allowing a company to proactively provide the best customer service possible while creating a database of customer preferences that can be reviewed by sales, marketing, and management. The data can then be used to reduce costs and improve employee productivity. (Bergeron, p. 1-2)

CRM is often touted as a means of improving customer satisfaction, which in turn improves long-term customer loyalty. It’s also ostensibly part of a public relations campaign to increase the company’s image. Many CRM companies claim that their products make it easier for employees to do their jobs; however, these subjective metrics are difficult to measure quantitatively. (Bergeron, p. 8)

Literature Review

Kubina and Lendel (2015) mentioned about the problems of successful application of social CRM in the company are still poorly explored and in practice weakly applied area of marketing management. Managers often do not know modern techniques and metrics for assessing implementation of social CRM. The aim of their paper is to create a system for successful application of social CRM in the company based on detailed analysis of sources and conducted research.

Harrigan, Soutar, Choudhury and Lowe (2015), tried modelling CRM in a social media age. They said that CRM is a continually evolving domain that has been particularly affected by social media, which have revolutionised the way businesses and consumers interact. Their paper on social CRM builds on a previous model of CRM prior to the growth of social media. They present a new model for social CRM, including a new construct of customer engagement initiatives and adaptations of other constructs to cater for the impact of social media. An online survey was used to collect data from a population of marketing practitioners and partial least squares analysis was used to test the model. Findings show the importance of an underlying customer relationship orientation; how it impacts on social media technology use and customer engagement initiatives, and also directly on customer
A relationship is also shown between engagement and relational information processes, which is viewed as a performance outcome of social CRM. Thus, from a managerial perspective, one recommendation they make is that organisations should utilise the rich customer information that is created with every customer engagement through social media to drive future marketing decisions.

Santouridis and Tsachtani, (2015) investigated the impact of CRM resources on CRM processes and focused on a customer life-cycle in the case of a Greek Bank. Contemporary economy, which is characterized by globalization, increasing competition and advances in communication and information technology, force companies to depart from traditional marketing doctrines and adopt a customer-centric approach by focusing on managing customer relationships. Relationship marketing is a marketing strategy enforcing companies to deal with enormous customer data management challenges, thus necessitating the deployment of supporting IT systems. This need has been covered to a great extent by the development of CRM systems. The aim of the present study is to investigate the impact of a company's CRM related human, organizational and technological resources on its CRM processes. A customer life-cycle based approach has been chosen. As a result of this, CRM processes have been mapped on the initiation, acquisition, regain, maintenance, retention, expansion and exit customer life-cycle stages. Field research was conducted by utilizing an interviewer-administered questionnaire, which was developed by adopting relevant work reported in literature. The case industry chosen was the Greek banking sector. The survey was conducted among all the employees of a Greek leading bank's 10 branches located in the region of Thessaly, who were involved in CRM processes. The final sample comprised 102 correctly answered questionnaires. Inter-item analysis was used to verify the scale's factors for internal consistency or reliability.

The Cronbach's alpha, which was calculated for each scale, ranged between 0.819 and 0.912. Regression analysis were then performed to examine the impact of CRM resources on customer life-cycle stages. The results highlighted that CRM organizational resources are the most vital, since they have a positive effect on the processes of all customer life-cycle stages. Moreover, human resources were proved to have a significant effect on the early (initiation, acquisition, regain) and middle (maintenance, retention) customer life-cycle stages, while their impact on the late stages (expansion, exit) is minimal. Finally, technological resources were shown to have a small influence on the middle customer life-cycle stages.
Garrido, Lockett and García-Morales (2014) focused on paving the way for CRM success. They investigated the mediating role of knowledge management and organizational commitment. CRM has become one of the most influential technologies in the world, and companies are increasingly implementing it to create value. However, despite significant investment in CRM technology infrastructure, empirical research offers inconsistent support for its positive impact on performance. This study develops and tests a research model analyzing the process through which CRM technology infrastructure translates into organizational performance, drawing on the resource-based view (RBV) and the knowledge-based view (KBV) of the firm. Based on an international sample of 125 hotels, the results suggest that organizational commitment and knowledge management fully mediate this process.

Gholami, Saman, Sharif and Zakuan (2015) concentrated their research on “A CRM Strategic Leadership Towards Sustainable Development in Student Relationship Management”: Sustainable development (SD) is not just obligating society members to fulfill their responsibilities, but can also generate significant competitive advantage. Only recently, organizations increasingly consider SD as a business strategy because it can deliver larger shareholder value and provide access to capital and making stronger performance over time which can improve shared value for both business and society. Considering the present atmosphere where competition is the most important factor in asserting oneself and exerting influence in the developmental process, CRM and the significant influence it has brought about in the worldwide scene can be seen as one of the most influential factors contributing to the economic sphere. Applying this concept in higher education academic as a specialist manpower center of education, has led to a new concept named Student Relationship Management (SRM). This research by employing the CRM, a strategic leadership towards sustainable development in SRM and subsequently in higher education, addressed to identification of SRM factors with exploratory and confirmatory factor analysis (FA) in order to implementation of the SRM system. Thus, after identifying the quality elements of SRM, using FA through SPSS and Amos software, the new scale was standardized. The identification phase led to find the six factors along with 20 items based on literature, FA and this result is helpful for Universiti Teknologi Malaysia on how SRM could be linked to CRM and an improved the system would assist to place sustainability in higher education.
Qi, Qu, and Zhou (2014) searched the answer of how the customer self-construal moderate CRM value creation chain. Most of the existing literature on CRM value chain creation has focused on the effect of customer satisfaction and customer loyalty on customer profitability. In contrast, little has been studied about the CRM value creation chain at individual customer level and the role of self-construal (i.e., independent self-construal and interdependent self-construal) in such a chain. This research aims to construct the chain from customer value to organization value (i.e., customer satisfaction → customer loyalty → patronage behavior) and investigate the moderating effect of self-construal. To test the hypotheses suggested by their conceptual framework, they collected 846 data points from China in the context of mobile data services. The results show that customer’s self-construal can moderate the relationship chain from customer satisfaction to customer loyalty to relationship maintenance and development. This implies firms should tailor their customer strategies based on different self-construal features.

Trainor, Andzulis, Rapp and Agnihotri (2014) studied on the social media technology usage and customer relationship performance. They did a capabilities-based examination of social CRM. Their study examines how social media technology usage and customer-centric management systems contribute to a firm-level capability of social CRM. Drawing from the literature in marketing, information systems, and strategic management, the first contribution of this study is the conceptualization and measurement of social CRM capability. The second key contribution is the examination of how social CRM capability is influenced by both customer-centric management systems and social media technologies. These two resources are found to have an interactive effect on the formation of a firm-level capability that is shown to positively relate to customer relationship performance. The study analyzes data from 308 organizations using a structural equation modeling approach.

Josiassen, Assaf and Cvelbar (2014) tried to find an answer of the question of whether all CRM dimensions affect firm performance. Successful firms often endeavor to assure competitive advantages through the relationships with their customers. Consequently, CRM has become of pivotal importance to many firms. This study investigates the effect of each CRM dimension on the performance of hotels. They found that in general hotels should aim to improve CRM capabilities because it has a positive effect on firm performance. Contrary to some previous assumptions, CRM investments did not result in positive performance. These findings are important as hotels strive to allocate resources to improve relationships with customers.
Khodakarami and Chan (2014) explored the role of CRM systems in customer knowledge creation. This study explores how CRM systems support customer knowledge creation processes, including socialization, externalization, combination and internalization. CRM systems are categorized as collaborative, operational and analytical. An analysis of CRM applications in three organizations reveals that analytical systems strongly support the combination process. Collaborative systems provide the greatest support for externalization. Operational systems facilitate socialization with customers, while collaborative systems are used for socialization within an organization. Collaborative and analytical systems both support the internalization process by providing learning opportunities. Three-way interactions among CRM systems, types of customer knowledge, and knowledge creation processes are explored.

Nettleton (2014) focused on CRM and analysis in commercial data mining. The area of CRM has attracted a lot of attention, and many businesses who are end users of IT solutions have spent considerable amounts of money on implementing CRM systems integrated to a greater or lesser extent with their operational and business processes. However, what should be kept in mind is that CRM is a basic, common sense idea that can be put into practice with nothing more than a spreadsheet and a modest database. This chapter introduces the reader to CRM in terms of recency, frequency, and latency of customer activity, and in terms of the client life cycle: capturing new clients, potentiating and retaining existing clients, and winning back ex-clients. The chapter then discusses the relation of data analysis to each of the CRM phases and considers customer satisfaction and integrated CRM systems. Next, it briefly describes the characteristics of commercial CRM software products, and finally, the chapter examines example screens and functionality from a simple CRM application.

Šebjan, Bobek and Tominc (2014) studied on organizational factors influencing effective use of CRM solutions. In this paper they discuss the issues of effective use of information solutions of customer relations management. The conceptual model is developed based on the Technology Acceptance Model (TAM). In the literature they can see that researchers often analyze the usefulness of CRM solutions at the level of individuals, and less at the level of organization and its orientations. These aspects are included in the model, proposed in this article, by the set of organizational factors which may influence the effective use of CRM solutions. Organizational factors are reflected in the three types of orientations – in process, technological and innovation orientation of organization. The conceptual model is
presented and organizational factors are discussed in more detail. The original value of their paper is in the constructed conceptual model that combines use of CRM solutions at the level of organization and critical organizational factors. The platform of the conceptual model enables its future empirical testing.

Ruivo, Mestre, Johansson and Oliveira (2014) defined the ERP and CRM integrative value. The value of IT adoption has been and still is a crucial question for the decision on IT adoption. In their paper they suggest a research model that aims at defining the integrative value of adoption of Enterprise Resource Planning (ERP) and CRM systems. The integrative value is described from the Resource Based View of the firm (RBV) and will be measured as impact on firm performance. The research model suggests six hypotheses that will be tested and analysed with data from a questionnaire among firms that have adopted both ERP and CRM systems in their organization. Due to the nature of the research model and the fact that it has not been tested in the past, the data analysis will be supported by Partial Least Squares (PLS). Their aim with this research project is that it will provide new knowledge on how integration between systems can positively influence value from IT investments, but also how different software such as ERP and CRM provides value to systems integration as well as process integration.

Mekkamol, Piewdang and Untachai (2013) modeled e-CRM for community tourism in Upper Northeastern Thailand. The paper is designed to provide a quantitative measure of e-CRM for community tourism in Upper Northeastern Thailand. The purpose of their study is to develop the e-CRM model in the upper Northeast of Thailand. Specifically, the objective of their study is to examine the validity and reliability of the four-factor model in e-CRM. The research mainly involves a survey design. It includes a pilot test using undergraduate business students at Udon Thani Rajabhat University for pretesting questionnaire items. In addition, this investigate into website character, website contact interactivity, shopping convenience, and care and service attributes necessitates uncovering variables of interest and this involves a large-scale field study. The data are collected via personal questionnaires from 447 samples. They include the customers of hospitality and tourism industries in 3 provinces such as Udon Thani, Nongkhai, Loei. Respondents are asked to rate, on a five-point Likert scale, their agreement or disagreement on the e-CRM attributes. LISREL program is used for data analysis since the proposed model is a simultaneous system of equations having latent constructs and multiple indicators. Quantitative data are analyzed by the statistical techniques, namely exploratory factor analysis and structural equation modeling. It is found
from the study that the effect of website character and shopping convenience, on care and service through website contact interactivity. The managerial implications are discussed.

Chuang and Lin (2013) studied the roles of infrastructure capability and customer orientation in enhancing customer-information quality in CRM systems with empirical evidence from Taiwan. The study adopts both a resource-based perspective that combines technology, human, and business resources to develop an infrastructure capability, and a strategic-positioning perspective that emphasizes customer orientation to examine customer information quality in CRM systems. Specifically, this study examines how firms bundle infrastructure capability and customer orientation to enhance the quality of customer information that enhances customer relationships and firm performance. The results of data gathered from 116 financial service firms in Taiwan suggest that the impact of quality on firm performance begins with infrastructure capability and customer orientation, and that the complementarity between these factors positively influences customer information quality. The results indicate that customer information quality positively affects customer relationship performance, which consequently leads to improvements in overall firm performance.

Duffy, Fearne, Hornibrook, Hutchinson and Reid (2013) worked on engaging suppliers in CRM and the role of justice in buyer–supplier relationships. Given the crucial role of suppliers in collaborative supply chains, it is surprising that little attention has been paid to the nature and management of supplier relationships in the implementation of a retailer's CRM strategy. To address this gap in the extant literature, the theory of organizational justice is used to explore the extent to which perceived fairness in buyer–supplier relationships supports or inhibits supplier engagement with the CRM process. The rationale is that suppliers who feel fairly treated by key retail customers are more likely to invest resources in the acquisition and use of data central to the retailer's CRM strategy. By empirically testing a conceptual model linking downstream CRM to upstream SRM, the results provide evidence to indicate that customer data use is significantly influenced by perceptions of fairness, particularly with respect to the distribution of rewards, and the transparency of decision-making processes. As a key criticism of CRM centers upon the failure of organizations to exploit the full potential of customer data, the results highlight the usefulness of understanding the relational linkages between buyers and suppliers and the consequential behavior of suppliers in terms of engagement with customer data vital to the success of retailers’ CRM strategies.
Bahari and Elayidom (2015) worked An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. CRM-data mining framework establishes close customer relationships and manages relationship between organizations and customers in today's advanced world of businesses. Data mining has gained popularity in various CRM applications in recent years and classification model is an important data mining technique useful in the field. The model is used to predict the behaviour of customers to enhance the decision-making processes for retaining valued customers. An efficient CRM-data mining framework is proposed in this paper and two classification models, Naïve Bayes and Neural Networks are studied to show that the accuracy of Neural Network is comparatively better.

1. Problem Statement

In data mining road, the first step is the statement of the problem. The purpose of the study should be clearly identified. In this study the profile identification of the thermal hotel customers is done to able to identify strategies to increase the customer satisfaction and the number of customers visiting the hotel. CRM strategies are developed according to the analysis’ results.

2. Data Preparation

2.1 Data Set

Data that are extracted from thermal hotels data in Marmara Region in Turkey database is analyzed in Spss Clementine tool. After the data is taken to the Clementine program the other important step in data preparation is the identification of the data variables. In Table 2 the variable types can be seen. The range value stands for quantitative variables, flag value stands for two types variables, set value stands for more than two types variables.

2.2 Variables in the data set

There are 16 variables and 1016 data. These are: Reservation date, Check in date, Check out date, Stay of night, Reservation Type, Type of the room (Big suit room- small suit room), Pension (full pension-half pension), Country, Number of adults, Number of children, Baby bed, Honeymoon, Payment type (cash- credit card), Payment channel (Agency-customer etc...), Customer type (frequent customer, normal customer), Market details, Age, Gender and Marital Status. (Table 1)
3. Selection of the proper model

Before the analysis to able to get better data, data set should be understood and described for this purpose.

Descriptive Statistics

Reservation number, check in and checkout date variables are taken out of the data with “filter” mode for the thought that these variables have no effect and importance on the analysis result. (Table 3) After the removal of the variables by using filter node, the Tables 4 and 5 are gathered.

In the table the data records can be explained like this:
Graph: The graphic demonstration of the table, Type: The variables type, Valid: The number of data valid in the analysis, Only for the range, Min, Max, Mean, Std dev, Skewness and unique is used to indicate how many data sources available with set and flag type variables are the other indicators to help the analyzer.

According to Table 4 “payment type” has 3 different options in the unique model. Age variable has a min value of “7 and max 78” with a mean of 49,076 and standard deviation 11,362.

By using the summary table the distribution of each individual variables could be seen. In the Thermal Hotel data set the distributions of the variables can be seen below.

It could be seen in Figure 1 that 74.8% of the customer who selected this thermal hotel is married and 25.2% are single.

The gender distribution of the customers are almost equal 50.1% women and 49.9% men.(Figure 2)

In the Figure 3 the age variable shows the normal distribution. The mean of the age is 49.076.

It could be seen in Figure 4 that the reservation are done more with the telephone with 74.21% percentage. The next is 14.37% e-mail, 6.1% fax and 5.31% hotel reservation.

With the 4 different option it is seen that 89.27% percentage standard room type are chosen. Little suit with 6.0%, big suit 2.76% and family room 1.97%. (Figure 5)

Customers prefer full pension with 68.41% percentage and Half pension 31.59%. (Figure 6)

Country

In the country distribution, although there are many different customers from different countries Turkish and the Arabic customers seems to choose this hotel most.
Turkey 70.77%
Arabia 14.96%
France 5.12%
Germany 2.07%
Russia 1.28%
Iran 1.18%
Italy 0.98%
Ireland 0.69%

In the hotel the number of single visitors are equal to the number of the couples. In the hotel people staying more than two visitors is not common.

Among the customers data having no child is common, rarely having one children can be seen in Figure 7.

Among the data of the hotel there is no honeymoon couple recorded % 100 stands for no. (Figure 8)

In the Figure 9 it can be seen that 67.32% of the payments are done in cash, 31.79% of the payments are done by credit card.

**Customer Type**

76.38% of the customers are visiting the hotel once. 23.62% of the customers are visiting the hotel more than once.

**Market details**

57.98% of the market experienced the hotel after the company leaded them, 33.55% came with the tourism agency and 4.56% is the firm employees.

**The Results of the Clustering Analysis**

Cluster analysis divides data into groups (clusters) that are meaningful, useful or both. Cluster analysis groups data objects based only on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar to one another and different from the objects in other groups. The greater the similarity within a group and the greater the difference between groups, the better or more distinct the clustering. (Tan, Steinbach, Kumar, p.490)

There are number of clustering techniques, but two of the most prominent are K-means and K-medoid. In this paper k-means algorithm is used. (Tan, Steinbach, Kumar, p.496-497)
In k-means algorithm first k initial centroids where k is a user specified parameter, namely, the number of clusters desired, are chosen. Each point is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the points assigned to the cluster. The assignment is repeated and update steps until no point changes clusters, or equivalently, until the centroids remain the same.

**Basic k-means algorithm**

1: Select k points as initial centroids.

2: repeat

3: Form k clusters by assigning each point to its closest centroid.

4: Recompute the centroid of each cluster.

5: until Centroids do not change.

After the variables distribution, to identify the customer profile, Clustering is chosen for the analysis. Marital status, gender, age, and country demographic variables found to be meaningful. After the determination of the variables k-means method is chosen and the cluster of the data set is executed. The number of the cluster is determined by the analyzer. In determining the cluster number the ones with no outliers and less differences should be chosen.

In this data set there is found to be 4 meaningful clusters. (Figure 10)

1. **Cluster:** Age mean 49 single man
2. **Cluster:** Age 47 married women
3. **Cluster:** Age 44 single women
4. **Cluster:** Age 51 married men

In the thermal hotel data set although “country” variable was in the analysis it has no effect in differentiation of the customers. To able to see the cluster profiles and the variables the visual data interface of the SPSS Clementine could be used.

**Results**

Customer satisfaction is so important for the companies in the tourism sector who have many competitors.

Taking preventive action for the possible failures in the service sector is also crucial. In this rough competition companies who are deaf to the customer requirements and not willing to increase their service level could be lost in the tourism market.
The communication between the sector and the customer starts before the vacation and it takes its last shape during the stay and if continues after the stay then it could be a long term relationship. For this aim a sophisticated customer profile should be gathered and should be tracked carefully.

The proposed models after the analysis can be used for the firm to know about themselves to make prediction about the clusters with unknown results and to guide for the decision making.

In strategy or decision making periods the proposed models will be used data should be frequently viewed and they should be up to date all the time. And new models should be done. Data gathered in this study could be used for increasing the customer satisfaction and be a guide for the CRM applications.

In the analysis it is seen that the reservations are done more with the telephone so the people who are doing reservations should be trained well and should have high skills in communication. Reservation with e-mail is highly seen with the e-mails seasonal information about the hotel should be given and the special days should be celebrated. Since every customer has a special reservation number by determining the chosen dates, season, and room type for each customer special treatment should be done. Four clusters gathered from the analysis should be studied carefully, and special studies should be done for each cluster.

In the data set 16 different variables belonging to the 1016 customers totally making 23368 different records exits. For the data mining solutions Clementine tool of the SPSS firm is used. Especially the ability of extracting data from different sources and the usage of the various models with one interface made this program preferable for our study. The output of the analysis shows that the service provided by the companies should be diversified due to the different preferences of the customers. This is seen more obvious in the tourism sector. For example the customer visiting the hotel once and a customer visiting the hotel frequently should be considered with a different perspective. By the help of this understanding service quality for different segment of customers could be increased. Also it might be helpful in the customer satisfaction and loyalty efforts in the company.
Tables

Table 1. Data Set

<table>
<thead>
<tr>
<th>Reservation Number</th>
<th>Check in date</th>
<th>Check out date</th>
<th>Stay in night</th>
<th>Reservation type</th>
<th>Type of room</th>
<th>Pension</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>10591</td>
<td>02.12.2010</td>
<td>31.12.2011</td>
<td>399</td>
<td>Hotel</td>
<td>Little suit</td>
<td>Full pension</td>
<td>Turkey</td>
</tr>
<tr>
<td>10591</td>
<td>02.12.2010</td>
<td>31.12.2011</td>
<td>399</td>
<td>Hotel</td>
<td>Little suit</td>
<td>Full pension</td>
<td>Turkey</td>
</tr>
<tr>
<td>11798</td>
<td>23.01.2011</td>
<td>03.02.2011</td>
<td>11</td>
<td>Standard room</td>
<td>Standard room</td>
<td>Full pension</td>
<td>France</td>
</tr>
<tr>
<td>11798</td>
<td>23.01.2011</td>
<td>03.02.2011</td>
<td>11</td>
<td>Standard room</td>
<td>Standard room</td>
<td>Full pension</td>
<td>France</td>
</tr>
<tr>
<td>11840</td>
<td>01.03.2011</td>
<td>03.02.2011</td>
<td>11</td>
<td>Standard room</td>
<td>Standard room</td>
<td>Full pension</td>
<td>France</td>
</tr>
</tbody>
</table>

Table 2. Variable Types and Values

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Values</th>
<th>Missing</th>
<th>Check</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reservation Number</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check in date</td>
<td>Date</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check out date</td>
<td>Date</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stay in night</td>
<td>Integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reservation type</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of room</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pension</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>String</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of adults</td>
<td>Integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of children</td>
<td>Integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily bed</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Filtering of the variables

- Reservation Number
- Check in date
- Check out date
- Stay in night
- Reservation type
- Type of room
- Pension
- Country
- Number of adults
- Number of children
- Daily bed
- Number of guests
- Number of children
Table 4. Summary Table-1

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Skewness</th>
<th>Unique</th>
<th>Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room no. Numb.</td>
<td>Range</td>
<td>10591.000</td>
<td>13231.000</td>
<td>12893.725</td>
<td>918.473</td>
<td>-2.171</td>
<td>--</td>
<td>100</td>
</tr>
<tr>
<td>Room no. type</td>
<td>Set</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>Type of room</td>
<td>Set</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Payment</td>
<td>Flag</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>Country</td>
<td>Set</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>Number of adults</td>
<td>Range</td>
<td>0.000</td>
<td>1.000</td>
<td>0.976</td>
<td>0.152</td>
<td>-0.276</td>
<td>--</td>
<td>100</td>
</tr>
<tr>
<td>Number of children</td>
<td>Range</td>
<td>0.000</td>
<td>1.000</td>
<td>0.033</td>
<td>0.130</td>
<td>0.196</td>
<td>--</td>
<td>100</td>
</tr>
<tr>
<td>Baby bed</td>
<td>Range</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>--</td>
<td>--</td>
<td>100</td>
</tr>
<tr>
<td>Hispanic</td>
<td>Flag</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Payment type</td>
<td>Set</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>3</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5. Summary Table-2

Figures

Figure 1. Distribution of the Marital Status

Figure 2. Distribution of the Gender
Figure 3. Histogram of Age

Figure 4. Distribution of Reservation Type

Figure 5. Distribution of Room Type

Figure 6. Distribution of Pension

Figure 7. Histogram of Number of Children
Figure 8. Distribution of Honeymoon

Figure 9. Distribution of Payment Type

Figure 10. Cluster Profiles

References